

09/623977

433 Rec'd PCT/PTO 11 SEP 2000

11/p.r.f.s

1

TELECONFERENCING SYSTEM

in 81

This invention relates to audio teleconferencing systems. These are systems in which three or more participants, each having a telephone connection, can participate in a multi-way discussion. The essential part of a teleconference system is called the conference "bridge", and is where the audio signals from all the participants are combined. Conference bridges presently function by receiving audio from each of the participants, appropriately mixing the audio signals, and then distributing the mixed signal to each of the participants. All signal processing is concentrated in the bridge, and the result is monaural (that is, there is a single sound channel). This arrangement is shown in Figure 1, which will be described in detail later. The principal drawback with such systems is that the audio quality is monophonic, generally poor, and it is very difficult to determine which participants are speaking at any one time, especially when the number of participants is large.

15 An example is given in European Patent Specification 0291470. This discloses an arrangement in which some of the input symbols are inverted in phase before combining them in the return channel thus allowing the cancellation, for each user, of his own voice.

in 81
20 According to the invention, there is provided a teleconferencing system comprising a conference bridge having a multichannel connection to each of a plurality of terminal equipments, and at least one terminal equipment having means to separately process each channel to provide a plurality of outputs, each output representing one of the other terminal equipments. By adopting this multichannel approach, the conference environment can be tailored to the operating needs and
25 circumstances of each individual by participants themselves.

30 Preferably the conference bridge comprises a concentrator, having means to identify the currently active input channels and to transmit only those active channels over the multichannel connection, together with control information identifying the transmitted channels. This reduces the capacity required by the multichannel connection. The control information identifying the active channels may be carried in a separate control channel, or as an overhead on the active subset of channels. In a preferred arrangement the channel representing a given terminal is excluded from the output provided to that terminal. This may be achieved by excluding that channel from the processing in the terminal equipment,

001100 "091100" 001100

ms
B3

Figure 2 illustrates a spatial audio teleconference system according to one embodiment of the invention;

Figure 3 illustrates a N-channel speech decoder used in the embodiment of

Figure 4 illustrates a N-Channel audio spatialiser used in the embodiment

Figure 5 illustrates a second embodiment of the invention;

Figure 6 illustrates how the invention may be used with conventional

Figure 7 illustrates a variant of the invention for use with a video

Figure 8 illustrates a voice switched concentrator which may be used in

ing

In the conventional system illustrated in Figure 1 the conference bridge

located in the exchange equipment 100 receives signals from the various customers' terminal equipments 10, (20, 30 not shown) in response to sounds detected by respective microphones 11, 21, 31 etc. These signals are transmitted over the telephone network (1), to the exchange 100 at which the bridge is established. Generally the signals will travel by way of a local exchange (not shown) in which the analogue signals are converted to digital form, usually employing linear companding such as "A law" (as used for example in Europe) or "mu-Law" (as used for example in the United States of America) for onward transmission to the bridge exchange 100. On arrival at the bridge exchange 100, the bridge passes each incoming signal 11, 21, 31 through a respective digital converter 111, 112, 113 to convert them from A Law to linear digital signals, and then passes the linear signals to a digital combiner 120 to generate a combined signal. This combined signal is re-converted to A law in a further digital converter

5 two-way connection between terminal equipments 10, 20.

10 and a digital audio downlink 5. The audio downlink comprises several channels
51, 52. ~~Participants with suitable terminal equipment can then process these~~
channels 51, 52 in various ways as will be described.

latency and unreliable data rates. However, over the next few years rapid improvements are envisaged in this technology and it is likely to become the preferred telecommunication method. Such systems would be ideally suited to implementing this invention. The latest internet type modems provide 56kbit/s downstream (links 4,5:), and up to 28.8kbit/s upstream (link 3). They are low cost and are commonly included in personal computer retail packages. Ideally a system should be able to work with all of the above, and also with standard analogue PSTN for use as a backup.

30 terminal equipment 10 contains a microphone 11 and loudspeaker system 12 as before. However, the loudspeaker system 12 is a spatialised system - that is, it has two or more channels to allow sounds to appear to emanate from different directions. This may take the form of stereophonic headphones, or a more complex system such as disclosed in United States Patents 5533129 (Gefvert), 5307415



5 The output from the microphone 11 is encoded by an encoder 13 forming part of the terminal equipment 10, and transmitted over the uplink 3 to the exchange equipment 100. Here it is combined with the other input channels 21, 31 from the other participants, terminals into a concentrator 230 which combines the various inputs into an audio signal having a smaller number of channels 51, 52
10 etc. These channels are transmitted over multiple-channel digital audio links 5 to the customer equipments 10, (20, 30) where they are first decoded by respective decoders 14, 24, 34 (Figure 3) and provided to a spatialiser 15 (Figure 4) for controlling the mixing of the channels to generate a spatialised signal in the speaker equipment 12.

15 The concentrator 230 selects from the input channels 11, 21, 31 those carrying useful information - typically those carrying speech - and passes only these channels over the return link 5. This reduces the amount of information to be carried. A control channel 4 carries data identifying which channels were selected. In the terminal equipment the spatialiser 15 uses data from the control channel to
20 identify which of the original sound channels 11, 21, 31 it is receiving, and on which of the "N" channels 51, 52 in the audio link each original channel is present, and constructs a spatialised signal using that information. The spatialised signal can be tailored to the individual customer, for example the number of talkers in the spatialised system, the customer's preferences as to where in the spatialised
25 system each participant is to seem to be located, and which channels to include.

In particular, the user may exclude the channel representing his own input 11, or may select a simultaneous translation instead of the original talker.

Transmission efficiency is achieved because only the active subset N of the total number of channels M are transmitted at any one time. The subset is chosen using a voice controlled dynamic channel allocation algorithm in the N:M concentrator 230. A possible implementation of this is shown in Figure 8. Each input channel 11,21,31 is monitored by a respective analyser 231, 232, 233. As shown for analyser 231, the signal is subjected to a speech detection and analysis process 231b. This detects whether speech is present on the respective input 11,

5

10

20

30

5
10

1

20

25

30

known as "transaural". As with ambisonic systems, the correct listening region is very small.

The output of several spatialisers may be combined as shown in Figure 4, which shows a spatialiser group for a stereophonic output having left and right channels 12L, 12R. Each channel 51, 52, 53 is fed to a respective spatialiser 151, 152, 153 which, under the control of a coefficient selector 150 control by the signals in the control channel 4, transmits an output 151L, 151R etc to each of a series of combiners 15L, 15R. The processing used to create the outputs 151L, 151R etc is operated under the control of the control signal 4 such that each channel appears as a virtual sound source, having its own location in the space around the listener.

The positions of virtual sources in three dimensional space could be determined automatically, or by manual control, with the user selecting the preferred positioning for each virtual sound source. For a video conference the positioning can be set to correspond with the appropriate video picture window. The video images may be sent by other means, or may be static images retrieved from local storage by the individual user.

If the spatialised sound is relayed via loudspeakers 12, rather than headphones, it will be necessary to prevent signals from the loudspeakers 12 being picked up by the microphone 11, re-transmitted and being heard as an echo at the distant sites 20, 30 etc. A technique for achieving this will be described later, with reference to Figure 11.

Figure 5 shows an alternative arrangement to that of Figure 4, in which the spatialisation is computed in the conference 'bridge'. Each conference participant receives the same spatialised signals, thus simplifying the customer equipment. Figure 5 is similar in general arrangement to Figure 2, except that the decoder 14 and spatialiser 15 are part of the exchange equipment 200. The output from the spatialiser 15 is passed to an encoder 18 which transmits the required number of audio channels (e.g. two for a stereo system) to each customer 10, 20, 30. This requires the number of channels in the downlink 5 to be equal to the number of audio channels in the spatialisation systems' outputs, instead of the number selected by the concentrator (plus the control channel 4) as in the embodiment of Figure 2. It also simplifies the customer equipment 10. However, this arrangement requires all customer installations 10, 20, 30 to have similar

5

10

30

20 The system described above with reference to Figure 4 employs linear artificial spatialisation techniques. Figure 11 shows how this, and the fact that the echo from each loudspeaker 12L, 12R combines linearly at each microphone 11L, (11R, not shown), allows echo cancellation to be provided for each output channel 3L, (3R) by having a separate adaptive filter 161L, 162L, 163L, (161R, 162R, 25 163R) on each input channel 51, 52, 53. Thus the adaptive filter 161L will model the combination of the spatialiser 151 for the channel 51, and the echo path between the loudspeakers 12L and 12R and the microphone 11L. This arrangement is discussed in detail in the applicant's co-pending application claiming the same priority as the present case.